Nanotechnology Informatics White Paper

Prepared for the National Cancer Institute by the caBIG® Integrative Cancer Research Nanotechnology Working Group

February 2009

Contributing authors (alphabetical):
Nathan A. Baker, Martin Fritts, Samira Guccione, David S. Paik, Rohit V. Pappu, Anil Patri, Daniel Rubin, Stanley Y. Shaw, Dennis G. Thomas

Table of Contents

Table of Figures

Acknowledgements

The caBIG® Integrative Cancer Research (ICR) Nanotechnology Informatics Working Group would like to thank Mark Adams, Elaine Freund, Juli Klemm, Li Kramer, Krzysztof Ptak, and Baris Suzek for the useful feedback.

Acronyms

BFO	Basic Formal Ontology
caBIG®	Cancer Biomedical Informatics Grid
caDSR	Cancer Data Standards Repository
CCNE	Center for Cancer Nanotechnology Excellence
CDE	Common Data Element
ChEBI	Chemical Entities of Biological Interest
EVS	Enterprise Vocabulary Services
FDA	Food and Drug Administration
GO	Gene Ontology
ICR	Integrative Cancer Research (Workspace)
MeSH	Medical Subject Headings
MIAME	Minimum Information About a Microarray Experiment
NCBO	National Center for Biomedical Ontology
NCI	National Cancer Institute
NCL	Nanotechnology Characterization Lab
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NP-CDT	Nanoparticle for Cancer Diagnostics and Therapeutics
NPO	Nanoparticle Ontology
QSAR	Quantitative Structure Activity Relationship
SAR	Structure Activity Relationship
SDO	Standards Development Organization
SMILES	Simplified Molecular Input Line Entry Specification
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms
SPIO	Super-paramagnetic Iron Oxide
VCDE	Vocabularies and Common Data Elements (Workspace)

1 Executive Summary

1.1 Purpose

The purpose of this white paper is to define a vision for an emerging interdisciplinary field of study that merges cancer nanotechnology and biomedical informatics. To date, these two major areas have had little overlap. Nanotechnology, broadly, has its roots in materials science and for many years has focused on non-biomedical applications and thus, there has been low visibility for nanotechnology in the field of biomedical informatics. Currently, there is a tremendous opportunity to bring computational methods and information science to cancer nanotechnology[1] that this paper will address and make recommendations toward.

As the informatics community becomes more engaged in this intersection, informaticists and computational scientists will gain a deeper understanding of the nature of information pertaining to nanotechnology, and in particular about nanomaterials and their physical, chemical and biological properties. While some analogies to existing informatics areas (e.g., cheminformatics) may apply, there are issues unique to nanotechnology that must be elucidated and addressed. A new informatics infrastructure is being and will be created to support cancer nanotechnology research, and much of the existing informatics infrastructure such as the National Cancer Institute's Cancer Biomedical Informatics Grid (caBIG®) will be adopted and adapted to this area.

As the cancer nanotechnology community becomes more aware of the potential of informatics, it will stimulate new ways of thinking about research tools. Among the low-hanging fruit is the potential to customize knowledge resources to cancer nanotechnology. Because nanotechnology involves a lengthy development pipeline from nanomaterials synthesis to physical characterization to *in vitro* and *in vivo* characterization to clinical applications, information integration will be key to addressing the needs of this community. Beyond that, dry bench computational methods, i.e., *in silico* methods, will begin to complement wet bench research, although an enormous amount of empirical data must be collected in order to facilitate these methods. This will be one of the central challenges for nanotechnology informatics.

1.2 Intended Audience

The intended audience for this white paper is a wide-reaching group. On the one hand, the paper is intended to encourage cross-fertilization between several different research communities such as the nanotechnology research community, the cancer research community, the biomedical informatics community, and the nanotechnology environmental, health and safety (EHS) community. On the other hand, this white paper is directed toward the relevant policy makers and other stakeholders who are looking toward the future of a more

[1] Throughout this manuscript, we refer to "cancer nanotechnology" specifically although in most instances, it could be replaced more broadly with "biomedical nanotechnology" or "nanomedicine".

computationally driven approach to nanotechnology research.

1.3 Summary

Numerous nanoparticles for cancer diagnostic and therapeutic applications are currently undergoing clinical trials (also referred to in this paper as Nanoparticles for Cancer Diagnostics and Therapeutics or NP-CDTs). However, characterizing and understanding the *in vivo* effects of these novel materials stands as a potential roadblock to their deployment. Thus, an informatics infrastructure that facilitates sharing of nanoparticle related information is critical. In fact, this informatics infrastructure could potentially address numerous aspects of the cancer nanotechnology development process as shown in Figure 1.

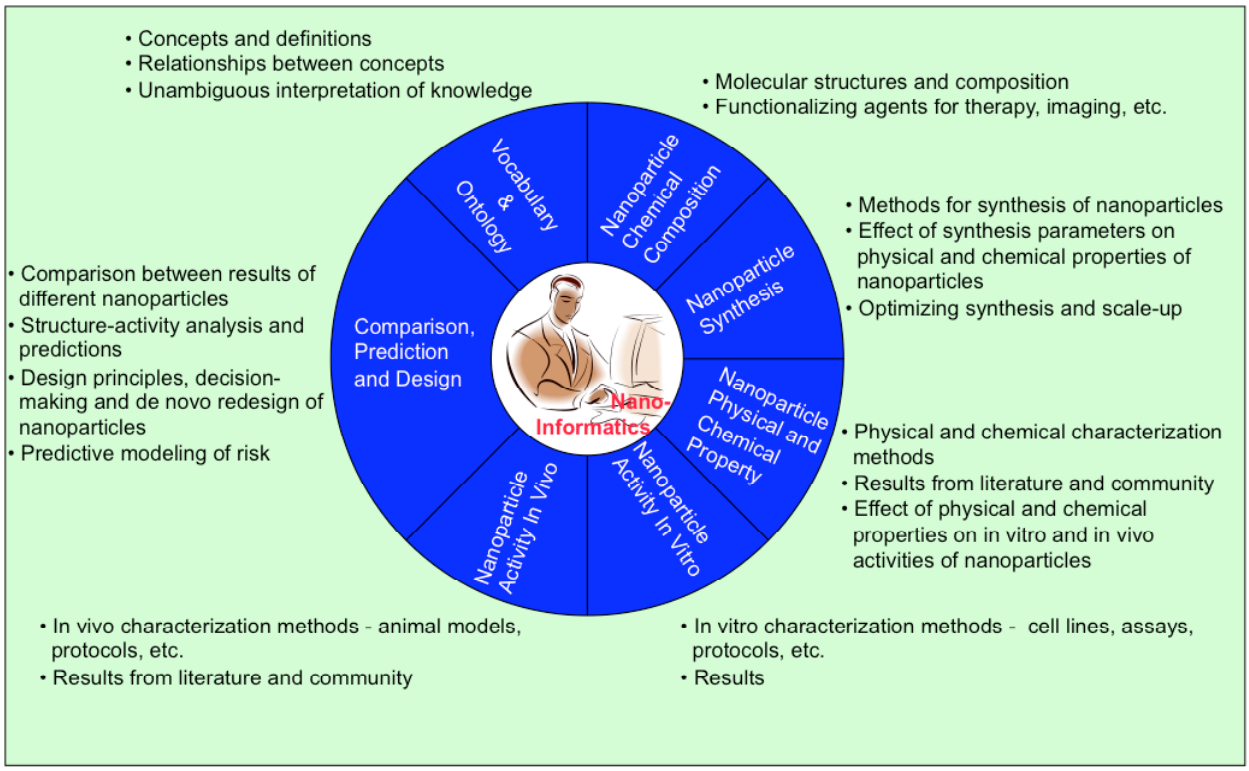


Figure 1: Making the case for a nanoinformatics resource in caBIG®: The figure depicts anticipated needs of users (shown as specific / directed queries outside the circle) that could be facilitated by the development of vocabularies and ontologies. The use of these vocabularies for data annotation, database creation and analytic tools will facilitate many applications including structure-activity relationships and clinical translation of nanotechnology.

There are numerous information needs of researchers, clinicians and the public that demand facile access to knowledge about cancer nanotechnology that is currently difficult to access and impedes progress. Although cancer nanoinformatics is in its earliest stages, there are several key resources under development including the caNanoLab database for storing/searching nanoparticle characterization data and the NanoParticle Ontology (NPO) for representing concepts within this space.

It is envisioned that computational methods will become indispensible in cancer

nanotechnology research. In order to advance in this direction, there must be progress in several key areas. First, the continued development and expansion of standardized vocabularies and ontologies is essential to enable machine- and human-interpretable comparisons of nanoparticle properties across numerous data sources. The success of the aforementioned NPO will depend on expanding its development and maintenance to a broader community and working toward widespread adoption as a standard within caBIG®. Second, a set of minimum information standards should be developed and adopted in order to uniformly provide critical characterization information that is frequently missing in the scientific literature. This should be done harmoniously with the continuing development of caNanoLab and similar to NPO, this minimum information standard should be developed, maintained and adopted by the community. Third, the scientific community should be encouraged to systematically generate dense matrices of data that will elucidate structure-activity relationships. A standardized suite of *in vitro* assays for biologically activity must be agreed upon, the data must be warehoused and an ethos of data sharing should be encouraged. Fourth, analytical tools and services should be developed that will complement the development of these critical information resources. Data analysis tools should be compatible with the numerous other tools within caBIG® and knowledge resources should be developed in order to create a portal for cancer nanotechnology information. And finally, the efforts of the cancer nanotechnology community must be coordinated and harmonized with the larger nanotechnology community, particularly the nanoEHS, regulatory and standards communities. Many different communities are currently developing their own vocabularies, definitions and ontologies with little effort to develop the framework for federating those ontologies into a system that allows integrated search and annotation across the many stove-piped databases. Tools such as BiomedGT could help to accelerate the harmonization of ontologies and the formation of a larger community of interest in nanobioinformatics. As illustrated in Figure 1, these informatics technologies have the potential to address numerous user needs and form the underlying technology and infrastructure that will enable quantitative structure activity relationship prediction.

2 Cancer Nanotechnology

2.1 History of Nanotechnology

Richard Feynman first introduced the idea of nanotechnology in 1959 with his oft-cited talk "There's Plenty of Room at the Bottom" that predicted the potential of manipulating systems on a small scale, down to the atomic level [1]. Today, nanotechnology is a major research and development growth area with the multi-agency National Nanotechnology Initiative (NNI) investing over $1.4 billion per year. Nanotechnology applications span a very wide spectrum of scientific research, including many very important life sciences applications [2]. Nanotechnology is a new and promising approach to cancer diagnostics and therapeutics with applications in imaging, early detection, therapeutic monitoring and multifunctional therapeutics.

The exponentially growing interest in nanotechnology in biomedical research is evidenced by the plot of number of publications per year shown in Figure 2. The growth of scientific publications on microarray technology is shown for comparison.

2.2 What makes nanotechnology useful?

Nanotechnology refers to novel materials synthesized via the control of matter on the nanometer (1-100 nm) length scale and the exploitation of novel properties and phenomena that emerge at this length scale [3]. In particular, nanomaterials can exhibit electrical, mechanical, optical, chemical and/or biological properties that are fundamentally different from those of the bulk material. For instance, the ability to modulate nanoparticle size and shape can dramatically affect pharmacokinetics [4, 5]. Nanoparticles can be decorated with multivalent delivery, imaging and/or therapeutic moieties, which increase avidity for the desired target, increase target-to-background ratios for imaging and optimize drug delivery [6-8]. Nanoscale materials can be also designed with very large surface areas and complex topologies [9, 10], further enhancing their ability to provide reactive surfaces, and deliver therapeutic payloads. Patterning of materials at nanometer scale can influence the wavelike properties of electrons in matter, allowing properties such as magnetization [11] and membrane penetration [12] to be fine-tuned while maintaining constant chemical composition.

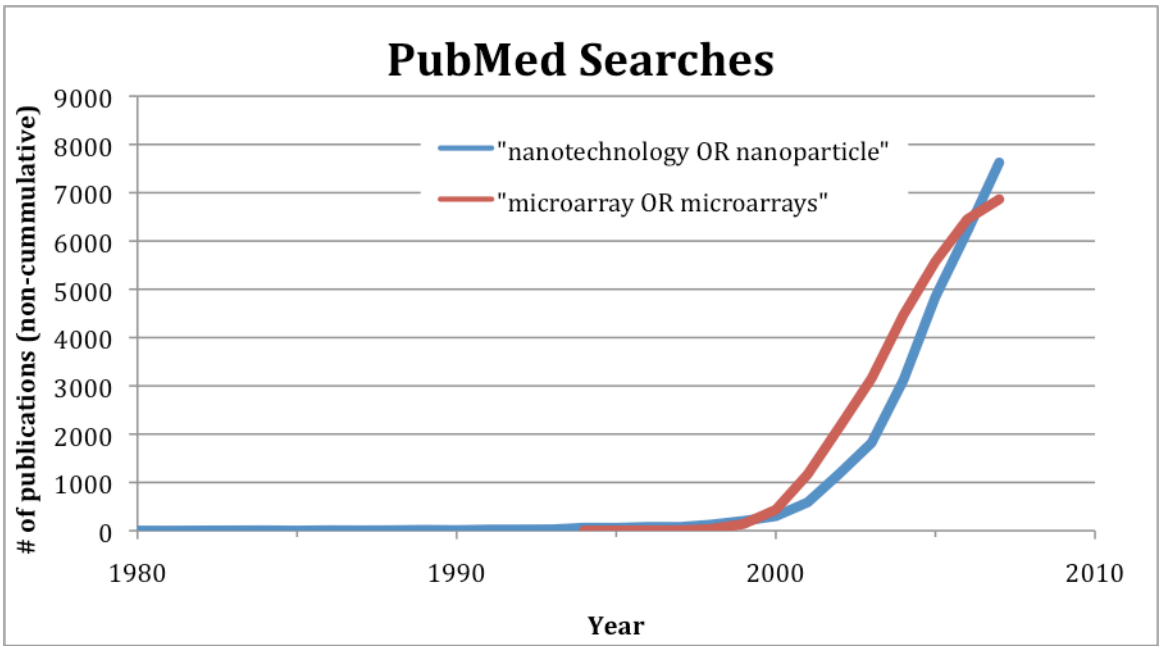


Figure 2: Comparison of the appearance of nanotechnology/nanoparticle in the biomedical literature as compared to microarray(s). Note that nanotechnology/nanoparticle is likely underrepresented here due to limited indexing of some physics, chemistry and materials science journals in PubMed.

2.3 Space of Nanoparticle Composition and of Physical/Chemical/ Biological Properties

The power of nanotechnology derives, in part, from the variety of nanomaterials that can be created by combinations of components: core constituent materials, surface coatings, and a wide variety of decorating moieties for targeting, imaging and therapy [13, 14]. These components can be combined using a variety of topologies and connectivities. Furthermore, the spectrum of sizes, shapes and charges of otherwise similar nanomaterials plays an important role in determining physical and biological properties. Thus, the space spanned by

potential nanomaterials is much larger and more complex than the range of conventional small molecule therapeutics, which themselves are potential nanoparticle payloads.

Similarly, the space of physical, chemical and biological properties of these nanomaterials is very large. We invoke the paradigm of a high-dimensional "caNP space" for NP-CDTs. Data can be parsed into three broad categories: data pertaining to chemical composition, data pertaining to physical properties, and data to quantify the biological functions of nanoparticles as diagnostic or therapeutic devices. Each complete data set constitutes a single "point" in a multidimensional caNP space, shown in Figure 3 as an effective 3-dimensional space. It should be stressed that a "point" along an axis is in reality a vector in a much higher-dimensional space. For instance, a point along the physical properties axis is characterized by multiple attributes including size, shape, flexibility, transport coefficients, hydrophilicity etc. Complete datasets for different NP-CDTs become unique points in this caNP space. In some cases, there will be overlap in terms of chemical composition and morphology while in some others there may be overlap either in physical or biological attributes such as the specificity of interactions with tumors or neovasculatures. In some other cases, the data are unlikely to be complete or comprehensive and the caNP space is only fractionally populated.

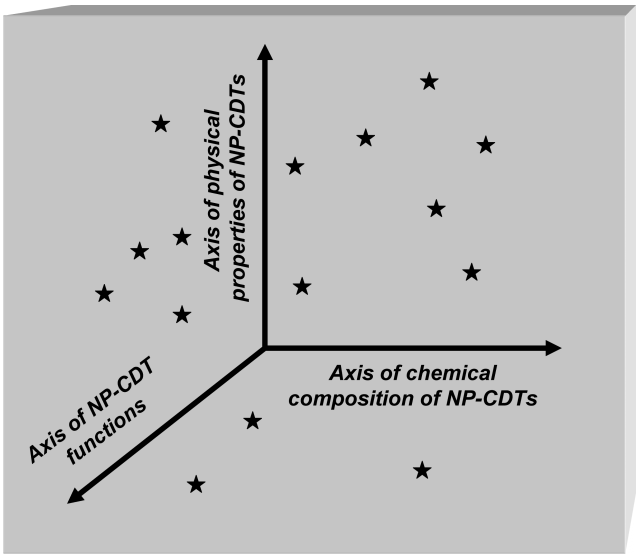


Figure 3: Depiction of the caNP space for NP-CDTs. Each "point", shown as a star, represents a complete data set.

2.4 Applications of Nanoscale Materials to Cancer and Other Human Diseases

Nanomaterials can provide significant advances throughout the spectrum of cancer diagnosis, treatment and monitoring. Nanomaterials can facilitate diagnosis [15] through new molecular imaging probes [16, 17], or nanosensors that can detect biomarkers in clinical specimens [18] and peripheral blood samples [19]. Nanoscale materials can impact therapy through several mechanisms [20], including improved drug delivery, pharmacokinetics and molecular targeting. Imaging and nanosensors can further provide mechanism-based treatment monitoring. Multimodal nanomaterials can combine both targeting and therapeutic moieties

[8, 20], and so called "theranostic" nanoparticles can carry both therapeutic and monitoring capabilities [7]. Taken together, nanoscale materials could greatly enhance a cycle of "personalized medicine" by elucidating a molecular phenotype or fingerprint of a patient's cancer, delivering molecular targeted therapies, monitoring the efficacy of treatment and informing future treatment decisions.

Some of these capabilities have already reached patients, either in the form of FDA-approved therapies or advanced clinical trials. For instance, superparamagnetic iron oxide (SPIO) nanoparticles are currently in clinical trials to noninvasively detect metastatic lymph node involvement in a variety of solid tumors [21]. Albumin-nanoparticles bearing paclitaxel (Abraxane®, Abraxis Bioscience, AstraZeneca) were FDA-approved in 2005 for patients with metastatic breast cancer [22]. A dendrimer-based microbicide (SPL7013, VivaGel®, StarPharma) is in Phase II trials for prevention of HIV and genital herpes [23]. A variety of nanoparticles are already in use for *in vitro* diagnostics, such as a gold nanoparticle-based system to detect polymorphisms in the CYP2C9 and VKORC1 genes, which affect metabolism and thus dosing of the anti-coagulant warfarin [24]. New devices that combine microfluidics and nanosensors are under investigation for detection of circulating tumor cells and biomarkers from peripheral blood [25, 26].

Another application for nanomaterials is the "rescue" of drug candidates that have failed clinical trials. Often, failure is attributed to insolubility, non-specific delivery and toxic side effects. With increased data sharing and openness fostered by caBIG® initiatives, it should be possible to develop trials to "rescue" failed drugs by loading them onto or encapsulating them into targeted nanoparticles. However, rational design of "rescue strategies" will need predictions obtained from data-driven models that have been tested and validated.

2.5 *Defining* in vivo *Effects: Critical for Clinical Translation of Nanotechnology*

Despite the successes of nanomaterials in humans and animal models, the fundamentally novel composition and behavior of these nanoscale materials could pose a translational roadblock to the ultimate application of these technologies to human disease and patient care [27]. For conventional small molecule or biologic therapies, the accumulated experience over decades can shed light on how certain functional units might behave *in vivo*. Even so, toxicity and adverse events can occur unpredictably during therapeutic development or following FDA approval. For nanoscale materials, there is significantly less understanding of how they interact with biological systems, and thus much less insight into *in vivo* structure-activity relationships. This uncertainty is compounded by the great variety of different nanomaterial platforms and compositions that are theoretically possible, by the interference of nanoparticles with standard assays, and by the need for improved separation technologies to better control the purity and batch-to-batch consistency of nanoparticle therapeutics and diagnostics. The concern over the potential risks of nanoscale materials developed to diagnose or treat cancer is a microcosm of broader societal uncertainty over the long-term consequences of exposure to nanomaterials. Indeed, hundreds of consumer products are already on the market that contain some component synthesized using nanoscale technologies [28]. This subject has been discussed at length in a recent report from the U.S. National Academies [29]. In a 2007 report on nanotechnology [30], the U.S. Food and Drug Administration (FDA) noted how nanotechnology could impact multiple areas under its regulatory purview, including new

therapies or devices, and raised several salient questions. Questions addressed in this report include:

- Are long-established surrogate assays for toxicity (e.g., DNA damage) appropriate for nanomaterials, or do new surrogates need to be developed and validated?
- Because a slight change in nanomaterial design can change its physical, chemical or biological properties, does every individual formulation need to undergo toxicity testing, or is some degree of extrapolation prudent?

These and other reports all cited the need for new methods of risk assessment on nanomaterials, including both *in vitro* and *in vivo* data, to integrate the efforts of scientists (academic and industry), clinicians, patients and regulatory bodies in bringing nanotechnology to the clinic.

2.6 Importance of Informatics to the Clinical Application of Nanotechnology

The safe application of nanoscale materials to human disease would certainly benefit from the types of quantitative structure-activity relationship (QSAR) models that exist for small molecule therapies. However, such QSAR models will require the collection of much more experimental data, since formulations containing nanoparticles are inherently polydisperse. However, since the nanocomponents of these formulations may exhibit different properties or interferences as a function of their small size, control over their polydispersity is much more important than for conventional small molecule therapies. Equally important is the ability to systematically collect, integrate and analyze data on large numbers of nanomaterials, tested in multiple *in vitro* and *in vivo* systems. This data collection and processing, in turn, requires collaboration and communication among stakeholders, as well as generalizable data standards, ontologies, and bioinformatics analytic approaches. In short, informatics approaches are critical to the translation of nanomaterial research into clinical and other applications.

Data-sharing across the nanomaterials research community is an important catalyst for these informatics approaches. Because of the uncertainty over how nanomaterials interact with *in vivo* biological systems, it is essential that data (including *in vivo* data from animal models and humans) is systematically collected and publicly accessible. This is true for not only successes, but perhaps more importantly, for "failures" (whether for reasons of pharmacokinetics, bioavailability, efficacy, or toxicity). However, a tendency to selectively publish positive results, and sequester other data in proprietary databases may decrease the ability of the community to learn about, and learn from, these negative findings. To help mitigate these issues, the nanotechnology community may benefit from the standards established for clinical trial data on conventional therapies. In 2004, the International Committee of Medical Journal Editors (ICMJE, representing the editors of 11 journals) mandated that clinical trials submitted for publication must be registered in a clinical trials registry at or before the time of subject enrollment [31]. In addition, the ICMJE outlined best practices that these registries should follow: free public access, management by a non-profit organization, the capability for electronic searches, and minimum information standards for study design and the interventions tested.

Spurred by this initiative, there are now several clinical trial registries that meet these criteria, including http://www.clinicaltrials.gov, a large clinical trial registry sponsored by the U.S.

National Library of Medicine[2]. The World Health Organization's (WHO) International Clinical Trials Registry Platform (ICTRP; http://www.who.int/ictrp) provides a platform to identify and classify trials contributed by a host of smaller registries, and represents a significant step towards the goal of a universal search portal for clinical trials. The systematic inclusion of nanomaterial studies into these databases may well require the establishment of distinct information standards, in order to capture nanomaterial composition faithfully.

The following sections review the current state-of-the-art and provide a gaps analysis for vocabularies and ontologies, standards, database resources and knowledge resources.

3 Vocabularies and Ontologies

Recent efforts have focused on the development of databases for storing cancer nanotechnology data, with the goal of catalyzing discovery by bringing together the key knowledge in nanotechnology into coalesced resources. One such resource is the caNanoLab project (http://gforge.nci.nih.gov/projects/calab/), which is being developed for the purpose of storing, searching, and sharing data generated from a variety of characterization studies of nanoparticles used in cancer-related research. However, to maximize utility, databases must be complemented by a common controlled vocabulary (or "controlled terminology") that can facilitate data sharing and cross-integration of cancer nanotechnology databases with each other and with related cancer-related databases. Existing vocabularies for concepts related to cancer biology (cancer cell lines, tumor types, animal models, genes, gene products, post translational modifications, and signaling pathways) and cancer medicine can be reused; e.g., NCI Thesaurus [32], Medical Subject Headings (MeSH) (http://www.nlm.nih.gov/mesh/), Gene Ontology (GO) [33], Chemical Entities of Biological Interest (ChEBI) [34], Pathway Ontology (http://rgd.mcw.edu/tools/ontology/ont_search.cgi), SNOMED_CT (http://www.ihtsdo.org/our-standards/snomed-ct/), etc. However, controlled vocabularies focused on cancer nanotechnology do not exist at this time. Developing vocabularies and ontologies in cancer nanotechnology will be essential to formally conceptualize, organize, and represent data in this domain so that that both humans and computers can meaningfully interpret and analyze them. While vocabularies are useful for providing standard names for the domain of discourse, the needs of cancer nanotechnology will be further enabled through development of formal representations of conceptualized knowledge through ontologies [35-37]. In biomedical research, ontologies are used to maintain the knowledge of a specific domain of interest in machine-processable form and to integrate experimental data that is annotated with concepts from these ontologies (see http://bioportal.bioontology.org), but there has been little focused effort on creating such ontologies in the domain of cancer nanotechnology. We anticipate that ontologies will be critical to cancer nanotechnology research due the intrinsically interdisciplinary nature of this research field. In particular, ontologies are needed to facilitate communication between cancer nanotechnology researchers from different scientific fields and to ensure semantic interoperability between applications and databases in this area of research.

[2] As of this writing, clinicaltrials.gov includes some 60 trials that involve nanomaterials that can be recovered by text searches.

3.1 The NanoParticle Ontology (NPO)

At Washington University in St. Louis, the Pappu and Baker research groups have been developing a NanoParticle Ontology (NPO) to represent the basic knowledge of physical, chemical, and functional characteristics of nanotechnology as used in cancer diagnosis and therapy. Initial versions of the NPO were based on a formal model for annotating nanoparticle formulations characterized in cancer research. This annotation model has been translated into an object model to provide a structural framework for annotating nanoparticle cancer research data using concepts from the NPO, controlled vocabularies, and other ontologies. Initial work on the NPO has provided a rich set of vocabulary relevant to nanoparticle formulations encountered in the literature, discussions with others in the cancer nanotechnology community, and curation of nanoparticle data from the scientific literature.

More recently, based on community feedback and internal review, the initial version of the NPO is being re-factored based on Basic Formal Ontology (BFO) principles for ontology design [38]. There are several reasons for this revision of the NPO. First, the original ontology had reached a design crisis: the annotation model that was originally used to guide ontology development led to ambiguous extension of the ontology during the introduction of new concepts. Second, original versions of NPO lacked the extensive Aristotelian logical structure that is necessary for future nanoparticle inference and design. The old structure of the NPO will remain as a "navigational" framework provided by the inferred structure of the ontology to assist users who are not comfortable with the stringent structure imposed by the asserted hierarchy established through the BFO-based design. However, the old structure of the NPO will not be used for primary growth and extension of the NPO.

Internal development of the NPO is performed using Protégé; public releases are available through the NCBO BioPortal (http://tinyurl.com/6ql696) and the BiomedGT Semantic MediaWiki (in progress). Figure 4 provides an overview of nanotechnology concepts from the initial version of the NPO and demonstrates the breadth of the current ontology. This ontology continues to grow through internal development, cancer nanotechnology data curation efforts, as well as community feedback through the NCI Alliance and caBIG® working groups. Although this ontology is under continual development and expansion, the current form provides over 490 classes for use in the annotation and exploration of nanoparticle data.

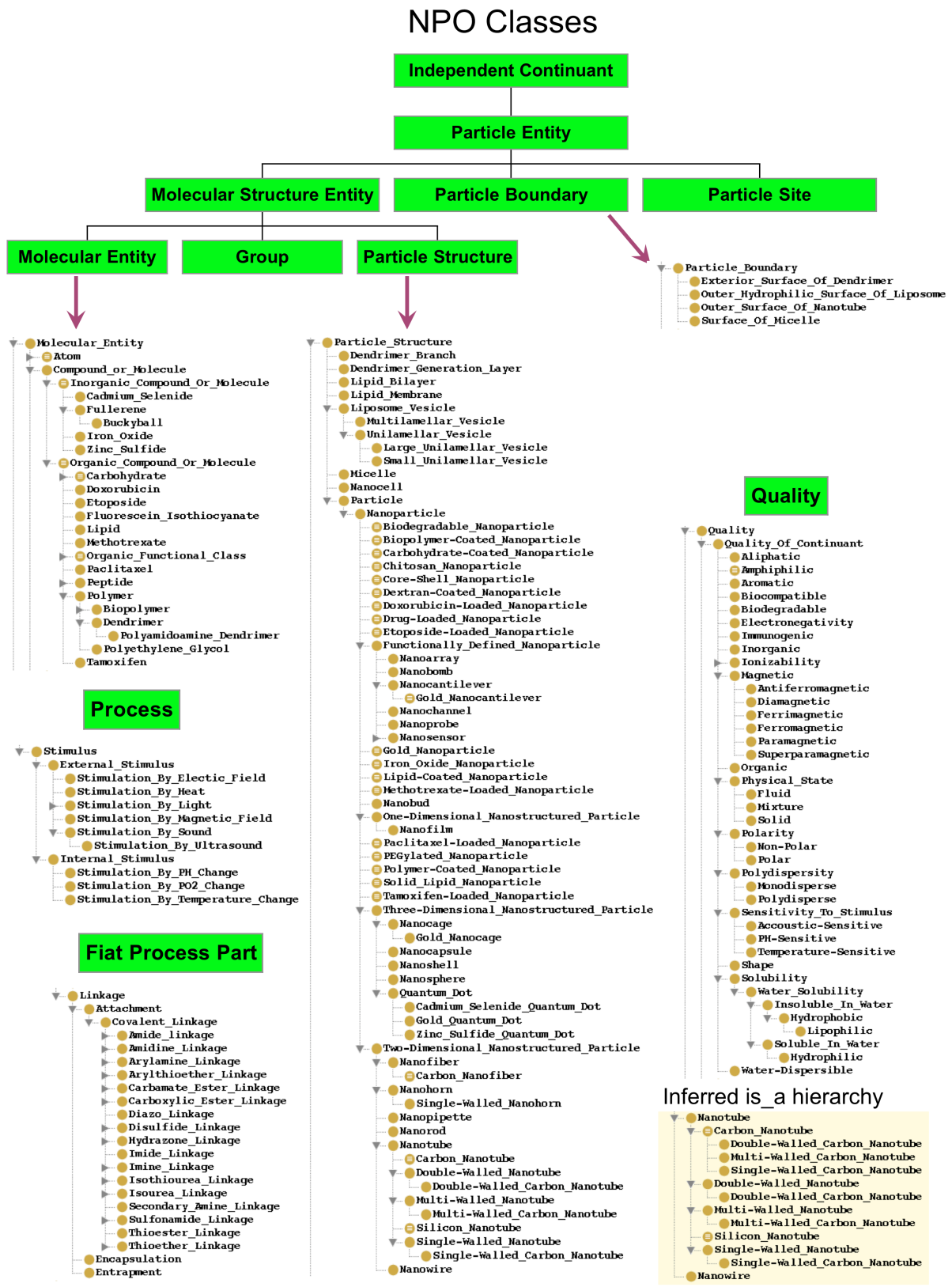


Figure 4: A subset of the top-level concepts provided by the NanoParticle Ontology (NPO).

3.2 Other Nanotechnology Vocabulary Activities

Although the NPO is almost certainly the most extensive ontology developed for nanotechnology, there are a number of significant efforts in vocabulary and ontology development for nanotechnology, most of which date from around 2000. The standards organizations ISO, ASTM, IEEE, and the OECD all have early stage efforts in ontology development as adjuncts to their projects in terminology definition, and the NNCO has been developing a high level ontology to serve as a core ontology for linking other ontologies. Most of these efforts resulted from briefings to these organizations on early versions of caNanoLab as well as the caBIG® framework and tools.

Two workshops were held in October 2008 that directly addressed the development and use of ontologies. The first, a Workshop on Enabling Standards for Nanomaterial Characterization, resulted in a consensus that a community of interest be formed to accelerate development of standard protocols and guides, reference material, inter-laboratory testing, and terminology and ontologies. The primary vehicle for accomplishing this goal is the establishment of the NanoCollaboratory, a wiki to facilitate communication among all interested US and international participants. The pilot wiki is currently now being populated and will be released in early 2009. The second workshop focused on increasing interoperability among the existing and planned databases supported by various US and international agencies and institutions. The primary result of that workshop was a consensus to initiate a demonstration portal to allow seamless data searches among caNanoLab, NanoHub, the Nano Manufacturing Network, databases at NIOSH and ONAMI, and the nanomaterial structures database and collaboratory developed by the Nano Linnaeus Project.

These workshops produced a broad consensus that a federated system of ontologies and vocabularies was needed in both standards development and nanobioinformatics. This consensus provides an opportunity to prototype and design such a federated system having both top-down participation from administrators of agency databases as well as bottom-up participation through industry, government, and academic participants through standards organizations. The NPO could be a major focus for this effort. Together with the caBIG® framework, the NPO could be further developed within BiomedGT, the standards wiki, and the new nano portal.

There are several other examples of ontologies for nanotechnology including an ontology for data-driven discovery of new nanomaterials [39, 40], a functional ontology [41, 42], the Nanotech Index Ontology [43], a proposed ontology for nanoscience [44], an example ontology of carbon nanotubes [45], a discussion of nanoinformatics with focus on ontology [46], and an atlas of nanotechnology [47].

Although the NPO is currently the most extensively developed nanotechnology-specific ontology, it is not the only source of nanotechnology vocabulary. In fact, terms relevant to cancer nanotechnology research can be found in a number of other vocabularies and ontologies, including NCI Thesaurus, ChEBI, Gene Ontology, SNOMED_CT, and many more. Such availability is not surprising, given the very interdisciplinary nature of nanotechnology and its application to biomedicine. The NPO provides a unique combination of new terms specific to nanotechnology as well as the organization of terminology present in other vocabularies to provide a logical description of cancer nanotechnology knowledge and data. The long-term goal is to harmonize NPO with existing terminologies pertinent to nanotechnology, so that NPO provides a comprehensive set of entities required to describe

nanotechnology data.

3.3 Opportunities for Ontology Use in Cancer Nanotechnology Research

As suggested above, ontologies will be very useful in enabling cancer nanotechnology research. The four key ways that ontologies can be used to enhance this work are as follows.

Defining the Key Entities and Relationships in Cancer Nanotechnology

The NPO as described above is a substantial initial step in establishing a comprehensive ontology in this domain. As emphasized in the NPO development process, initial focus must be on creating a comprehensive set of entities that will serve as descriptors to be used for annotating nanotechnology data. Human-readable (and subsequently, computer-interpretable) definitions of each entity are also vital at the outset. In the future, rich relationships can be added so that more complex computer processing and inference on these data will be possible.

Annotation of Nanotechnology Data

"Annotation" refers to the process of associating raw data or entries in a database with ontology terms for purposes of summarization, indexing, integration, and retrieval. Annotation is a key activity in many databases maintained by the Model Organism Databases (MODs). In many ways, the nanotechnology community has the same strengths and challenges as the model organism communities, thus adopting the MODs' methods will enable nanotechnology science as it has in biological science. Specifically, the challenges of integrating and accessing diverse data in nanotechnology as described above can be address by annotating these data using ontologies. If there was a consistent set of federated vocabularies and ontologies for nanotechnology, XML-enabled spreadsheets and documents could be incorporate those terms as tags (metadata) within the document or as row and column headings. This would provide an easy solution for uploading data, results, analyses, as well as annotations that would be searchable semantically while allowing full flexibility to the authors.

Data Integration

Data on the biological effects of nanomaterials are currently maintained in largely distinct databases. Structure-activity relationships are largely restricted to series of nanomaterials that utilize the same platform, and there is little capability to compare, for instance, a dendrimer and a quantum dot that are of similar diameter and bear similar effector or targeting moieties. As growing numbers of nanomaterials are synthesized and advance into clinical trials, the ontology of nanomaterials will facilitate the integration of a wide variety of data, from physico-chemical properties to *in vivo* activity and toxicity, and help identify the most promising nanomaterial properties for clinical application.

Enlarging the NPO to include structural models of the nanoparticles would permit the exchange of these models for collaborative computer modeling on physical, chemical, and biological properties of nanoparticles. The Nano Linnaeus Project database has already released a pilot version of this database [48]. An extension of the NPO to include modeling technique, as suggested as well by Hunter and others above, might also facilitate the new NCI effort to incorporate a broader spectrum of physical modeling techniques into cancer research and clinical practice. Care would have to taken to ensure that computational models of nanoparticle structures are not mistaken for experimentally determined structures if they were available.

Enabling Searching and Browsing for Nanotechnology Data

The ultimate goal of using ontologies to describe and annotate nanotechnology data is to enable discovery. Toward this end, researchers need to search for pertinent data sets or for specific knowledge about nanotechnology. Ontologies can enable search and discovery at the simplest level through indexing: finding records annotated with particular terms. However, the taxonomic structure of the ontology can also be exploited to expand and narrow searches along is-a, part-of, and other relationships, thereby allowing users to include related concepts in their search without the need for detailed knowledge of the terminology particular to a specific area of cancer nanotechnology research. Such semantic search capability can substantially reduce barriers to searching data across disciplines—an area of significant need in cancer nanotechnology research. For instance, synonymous search terms would be able to return the same results and related terms (by is-a or part-of relationship) would be able to be returned as well. This is discussed further in Section 6.

3.4 Current Vocabulary Development Problems and Needs

Nanotechnology vocabulary development currently faces a number of challenges, which will ultimately impact the usefulness of the vocabulary to the cancer nanotechnology community. The first and most obvious of these challenges has been met: the community has developed an initial "production" version of the NPO for public use. The following sections outline additional issues, which will need to be addressed by nanotechnology vocabulary developers.

Community Participation in Ontology Development

Currently, the core framework of the NPO is developed by a small group of researchers with specific interests in rational nanoparticle design. However, a broader range of cancer nanotechnology researchers must be engaged in future development and expansion of the ontology to ensure that the resulting terminology is valuable to the whole community. In addition, collaborative ontology editing tools are critical to these efforts. The NCBO BioPortal provides tools for ontology display and community feedback, which could be highly useful for this purpose. In addition, Protégé, an ontology editor, is being extended with new features to support collaborative development, and wiki-based methods such as BiomedGT are being developed. These tools should be evaluated as possible solutions to meet the needs of community collaborative ontology development in cancer nanotechnology. Ultimately, a broader support and governance structure will need to be established to ensure stable growth of the ontology based on community recommendations and expert review.

Relationships between Vocabularies

As mentioned above, there is strong overlap between many terms in the NPO and other established biomedical and chemical vocabularies. One significant challenge for continued development will be to harmonize the components of these vocabularies pertinent to nanotechnology to facilitate browsing, searching, annotation, and updates. This challenge is not unique to cancer nanotechnology vocabulary research but is, perhaps, exacerbated by the high degree of interdisciplinary work in this field. NCBO is developing tools that may be useful in this task; BioPortal provides means to declare mappings between terms across ontologies, and a means to store metadata in those mappings. Thus, it will be possible to create links establishing the fact that terms in different ontologies are synonymous, or related by homology across species, etc. However, significant curation work will be needed to realize such mapping and harmonization between vocabularies. The amount of effort might be

reduced dramatically by partially automating the harmonization. However, method development for ontology harmonization is an active research area and informatics and work is still underway to find efficient and robust methods for fully automatic ontology harmonization.

Adoption and Interaction with Users and Applications

The ultimate challenge for nanotechnology vocabularies is their adoption by the research community through use in databases, analysis applications, and search engines. There are several initial applications that could significantly benefit from nanotechnology ontologies and will serve as test cases for use of the NPO and related vocabularies as highlighted above. Many of these applications are best performed in the context of the caBIG® framework. NPO must become a caBIG® vocabulary standard in order to ensure that the ontology development efforts outlined above are useful to such caBIG® applications. Development of NPO as a caBIG® vocabulary standard will provide the basis for semantic interoperability in existing and emerging caBIG® applications for cancer nanotechnology research and will also help ensure common terminology usage as nanotechnology concepts are used in other research areas.

4 Nanotechnology Standards

One goal of nanotechnology efforts within caBIG® is to make nanomaterial data widely accessible and interpretable and thus, speed the development of novel diagnostic and therapeutic advances against cancer. This goal requires data that is produced using standardized protocols with appropriate controls to enable comparison between the results. The concept of standards is closely related to the data annotation applications discussed above. Generally, annotation should capture both the context (how the experiment was done) and the results (what was observed). In other biomedical communities, these annotation decisions are made by specifying the minimum information needed to describe experiments. The example of gene expression data is illustrative. A common data standard for gene expression data (Minimum Information About a Microarray Experiment, or MIAME) [49] enables investigators at multiple institutions to combine publicly available expression datasets and increase the power of their analyses. At the same time, annotation using the Gene Ontology (GO) classification systems for gene product attributes [33] (itself an effort to standardize gene product descriptions across multiple organism-specific databases; http://www.geneontology.org) enables mechanistic hypotheses to emerge from these dataset analyses.

4.1 Existing Nanomaterial and Protocol Standards Development Efforts

There are existing efforts at both nanomaterial and protocols standardization, such as the efforts at International Organization for Standardization Technical Committee on Nanotechnologies (ISO/TC 229), American Society for Testing and Materials (ASTM) E56 committee on Nanotechnology and American National Standards Institute's Nanotechnology Standards Panel (ANSI-NSP). ISO/TC 229 and ASTM E56 have ongoing efforts in the areas of terminology and nomenclature, measurement and characterization, and health/safety/environmental standards, all of which may provide opportunities for productive exchanges. The standardization efforts seek to facilitate the incorporation of nanotechnology

into commercial products, and eliminate technical barriers to global trade, which is true of virtually all standards organizations and the OECD.

This nanomaterial standardization process will facilitate and accelerate the development of technologies and provide a pathway for regulatory review and approval. Many classes of nanomaterial products are being developed as diagnostic devices for early detection, therapeutic and imaging applications. Beyond preclinical characterization, they must undergo regulatory review at the U.S. Food and Drug Administration (FDA) before clinical trials and eventual commercialization. The efforts at ASTM and ISO for standardization of protocols for characterization would facilitate the regulatory review process. The inter-agency collaborative efforts between the National Cancer Institute (NCI), the National Institute of Standards and Technology (NIST), and the FDA facilitate standards development process and help accelerate the translation of promising nanotechnology concepts to clinic. One of the goals of these collaborations is to develop Standard Reference Material (SRM) and protocol standards. With funding from NCI, NIST has already produced gold reference material standards (RM) at 10, 30, and 60 nm nominal size standards. These reference materials have been utilized in an inter-laboratory study (ILS) for characterization using standardized protocols developed at NIST and the Nanotechnology Characterization Laboratory (NCL). The results from this study will help compare results from different participating labs and address issues associated with sample preparation and protocol development. Such comparison will greatly facilitate and accelerate successful consensus standards development. Indeed, the IANH, the International Alliance for NanoEHS Harmonization, was recently organized specifically to promote formal and informal testing to accelerate development for nanotechnology. The development of the new NanoCollaboratory wiki will greatly accelerate the standards development process by providing an electronic means to develop standards collaboratively, track and archive all comments to reduce redundancy in discussion, arrange informal testing during protocol development, provide a single focus for standards development for nanotechnology to best leverage scarce resources and expertise, permit informal electronic ballot, and allow a seamless transfer of documented discussion to the target standards development organization (SDO). The same mechanisms will be available or future terminology development efforts by the SDOs and can be augmented by the use of semantic wikis and BiomedGT.

4.2 Unmet Needs in Cancer Nanotechnology Standards Development

These existing nanotechnology standards development efforts provide a good starting point for general nanotechnology concepts. However, the cancer nanotechnology community could greatly benefit from standards of minimum information specifically tailored to basic and clinical nanotechnology research.

Minimum information standards (e.g., MIAME) have become essential for allowing data from diverse laboratories to be collected and shared in such a way as to allow the understanding of the context, methods and data of an experiment [49]. In November 2008, the Minimum Information for Nanomaterial Characterization (MINChar) Initiative formed to create a minimum information standard for nanotoxicology studies. See http://characterizationmatters.org/. We propose that such a nanocharacterization standard should be collaboratively formulated and developed within the cancer nanotechnology community and harmonized with this effort in nanotoxicology. A minimum information standard for cancer nanotechnology research would describe the essential aspects of a

nanoparticle's physical, chemical, and biological characteristics, including:

- Nanoparticle synthesis, composition, and chemical properties
- Physical properties, including spatial, electrical, magnetic, optical, etc. characteristics
- Attached functional groups, e.g., for targeting, therapy, or imaging
- Biological properties, including pharmacokinetics, biodistribution, efficacy *in vitro* and *in vivo,* toxicity

The standard could also provide a set of optimal assays that can be applied within and between nanoparticle platforms and types.

Minimum Information for Biological and Biomedical Investigations (MIBBI) is a project (http://www.mibbi.org/) that maintains a large number of life sciences related minimum information standards and offers a 'portal' to provide visibility and access as well as a 'foundry' for modularization and community development. Re-use of available modules will prevent reinventing existing resources.

Taylor et al. have examined the commonalities between 21 checklist standards in MIBBI and found significant overlaps in study inputs, study procedures, assay inputs, assay procedures, data analysis, data description, data availability and administrative concepts. A standard for nanoparticle experiments would very likely find similarly high levels of overlap since nanoparticle experiments use a wide range of platforms already covered by other standards (e.g., cellular assays, gene expression, toxicity, proteomics, flow cytometry, molecular interactions and others).

Collaboration among developers of these nanotechnology standards and the community could promote awareness and usage of the standards. For instance, acquiring and inputting the required characteristics, metadata and data on each nanomaterial should not be excessively onerous. A governance and funding structure will need to be established (potentially involving a consortium of government, university and private groups) to ensure the long term maintenance of these standards. Finally, journal editors must feel that the standard is designed well enough to allow them to possibly use it as basis for required data submissions that would accompany publications. Currently, no such data submission policies exist for publications involving nanotechnology, and the research community would benefit from deposition of data on both positive and negative studies.

5 Database Resources

5.1 caNanoLab

Currently, the primary source of structured information for nanotechnology data is the caNanoLab software that was developed to support cancer nanotechnology research. caNanoLab (http://cananolab.abcc.ncifcrf.gov/caNanoLab/) is a caBIG® application designed for the curation and dissemination of nanoparticle physical, chemical, and biological data. Researchers at SAIC developed this database, under contract with the NCI Center for Biomedical Informatics and Information Technology (CBIIT), with feedback from users throughout the NCI Nanotechnology Alliance. caNanoLab is currently used by the Nanotechnology Characterization Lab (NCL) as well as the Washington University, Stanford University, and Emory/Georgia Tech Centers for Cancer Nanotechnology Excellence

(CCNEs). The database has been populated with data on over 270 different nanoparticles[3] from various participating laboratories and is experiencing rapid growth at the rate of 10-20 new nanoparticles per month. Other entities in caNanoLab include free-text reports on nanoparticles in PDF format as well as laboratory protocols. The rapid rate of caNanoLab growth has been driven by the use of this database for specific research needs by the NCL and CCNEs as well as a dedicated nanoparticle curator working at Washington University and supported by NCI/SAIC. This curator prioritizes, collects, and enters literature data on cancer nanotechnology research and also assists CCNE researchers with the dissemination of their data via caNanoLab. Such curation is very important to the continued growth of this resource because it (1) ensures the development of a strong base of nanoparticle data for subsequent mining and analysis, (2) broadens the data included in caNanoLab beyond the research areas of the core users, and (3) facilitates the use of caNanoLab for data dissemination by non-experts.

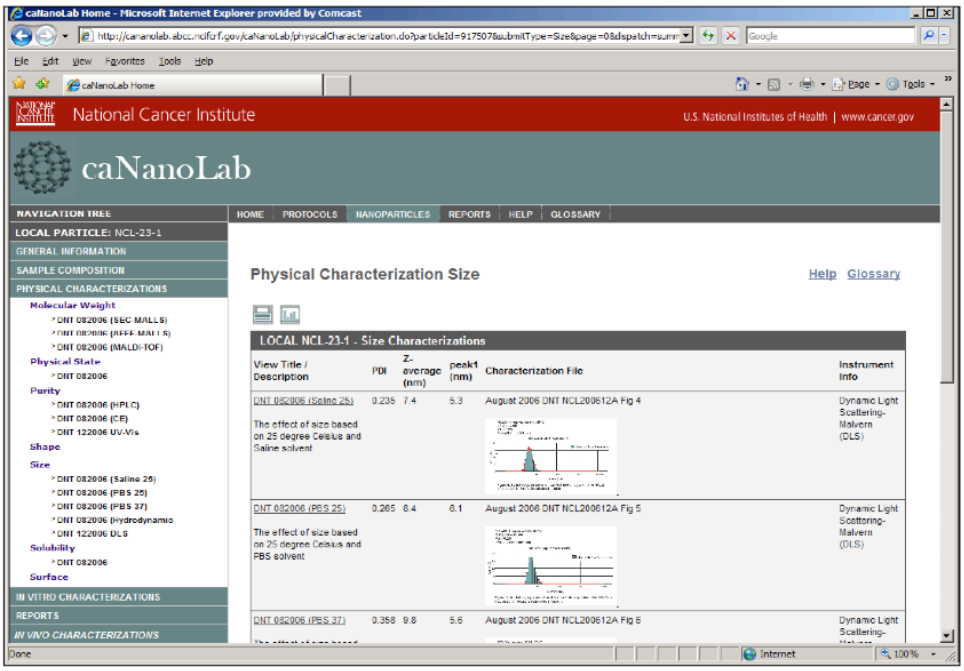


Figure 5: Using caNanoLab to search for nanoparticle physical characterization data.

caNanoLab has a web interface that uses a J2EE-based architecture (see Figure 5). However, as with most caBIG® applications, caNanoLab can also be accessed via caGrid so that the stored characterizations may be widely accessed by a federated approach. caNanoLab data structures are represented by the "nano" object model (nanoOM). The concepts associated with the caNanoLab object model are defined in the caBIG® Enterprise Vocabulary Systems, making them amenable to additional modes of query and browsing. Additionally, the metadata used by caNanoLab objects are maintained in the caBIG® data services registry (caDSR), which allows caBIG® developers to reuse common data elements and metadata between

[3] As of 27-Oct-2008, caNanoLab installations across the caGrid currently have a total of 274 nanoparticle entries and 50 protocols and documents available for public browsing.

applications. This is essential for interpreting the results of queries to caBIG® applications and for other data aggregation and analysis activities.

5.2 Non-nanotechnology Specific Databases

The goal of this section is to describe some of the resources and methods that are being developed in the broader caBIG® community that will likely be useful to nanotechnology research. There are many projects in caBIG® that could be relevant to nanotechnology. In the interest of succinctness, we will not discuss all of the activities, but will instead focus on imaging, as much nanotechnology data involves imaging. However, it should be noted that the techniques involved apply to non-imaging aspects of nanotechnology.

Image Storage and Repositories

The National Cancer Imaging Archive (NCIA) is a central repository of image from clinical oncology studies. These images are DICOM; metadata associated with the images are stored in XML files. The NCIA provides a Web-based front end for searching for images. This resource is very relevant to cancer nanotechnology since images are key to this domain. However, there are some challenges in directly adopting the NCIA infrastructure for nanotechnology.

1. **Non-DICOM image formats**: In nanotechnology, most of the imaging is in small-animals, and the images are in non-DICOM formats. NCIA, on the other hand, currently handles only DICOM. Thus, NCIA cannot be easily adopted by the nanotechnology community. There are several possible options: (1) DICOM wrapper objects for other image formats could be created. DICOM already has such a wrapper described in the standard; (2) a non-DICOM image archive could be developed as a caBIG® project; (3) non-DICOM images could be published on the Web and referenced by existing nanotechnology resources by URL. Due to variations in research workflow, creating a non-DICOM image storage solution is likely the best approach.
2. **Image metadata**: NCIA is in process of migrating image metadata to AIM format. The AIM format was designed to meet the needs of clinical researchers. While in concept AIM will be useful to the nanotechnology community, AIM would need to be extended to capture the rich image metadata associated with the nanotechnology domain.
3. **Image transmission and sharing**: Nearly all nanotechnology laboratories have internal image management solutions, and electronic data sharing is not a common paradigm. Beyond the social obstacles and intellectual property issues hindering data sharing, there are no standards or tools currently defined for labs to use for image sharing. It is possible that an existing solution currently used in clinical scenarios could be adopted by nanotechnology laboratories, such as the Clinical Trial Processor (CTP) of RSNA.

Data Resources

caBIG® contains several data resources (repositories of non-image data) for translational research, such as caArray, caTissue, and Rembrandt. They provide standards-based methods for describing the content of these databases (particularly by adopting terminology standards, interoperable data models, and APIs to the data enabling access across caGrid. There are clear synergies of these efforts with nanotechnology efforts, and the latter domain contains a wealth of data of varying data types that need to be described using standard terms, integrated, and

queried.

A key task that needs to be undertaken is an inventory of existing resources in nanotechnology and a list of the key data elements that these resources contain. That list of data elements can then be harmonized with elements (Common Data Elements—CDEs) of caBIG®. This will enable semantic interoperability among the existing caBIG® resources and that of nanotechnology projects undertaken as part of caBIG®. Much of the terminology relating to these data elements in nanotechnology likely already has been captured by the NPO as described in Sec. 3.1 above.

Need for Annotation (AIM)

A key aspect of using standard terminologies in caBIG® is their application to data sets in a process called "annotation." Annotation is particularly important in the Imaging Workspace, where a standard called Annotation and Markup (AIM) has been developed to address the challenge of making medical image content computationally accessible. AIM adopts knowledge representations for what people viewing images want to say about them: the entities observed in images (anatomy and abnormalities), the annotation contexts and image annotation requirements in those contexts to ensure the proper information is collected in the different contexts, and an annotation tool to create the annotations. AIM is a project of the caBIG® Imaging Workspace and is being developed to establish standards for recording semantic image information that will enable users to interoperate with these data nationally.

AIM distinguishes between image annotation and markup. Image annotations are explanatory or descriptive information, generated by humans or machines, directly related to the content of a referenced image (generally non-graphical, such as abnormalities seen in images and their locations). Image markup refers to graphical symbols that are associated with an image and optionally with one or more annotations of that same image. Accordingly, the key information content about an image lies in the annotation; the markup is simply a graphical presentation of the information in the annotation.

The AIM project provides methods for representing and handling both image annotations and markups. The approach to making the semantics of image content explicit and accessible to machines is to: (1) create an ontology to provide controlled terminology for describing the contents of medical images, and a standard image information model for recording semantic annotations, (2) develop an image annotation tool to collect user annotations as instances of the ontology, providing intelligent feedback to inform the user about annotation information requirements given the image annotation context, and (3) serialize the annotation instance data to a variety of standard formats for interoperability in many different imaging environments.

AIM provides an ontology in OWL-DL to represent the entities associated with medical images. The AIM ontology includes anatomic structures that can been seen in images (such as "liver" and "lung"), the observations made by radiologists about images (such as "opacity" and "density" of structures contained in the images), the spatial regions that can be visualized in images, as well as other image metadata (see Figure 6). The AIM ontology also represents knowledge about annotation requirements: information required to create image annotations. These annotation requirements are analogous to minimum information requirements in annotation tasks in other domains, such as in the microarray community. Annotation requirements comprise two aspects: (1) the context for the annotation, and (2) the requirement for annotation. The contexts for annotations comprise a set of pre-enumerated types of images and scenarios in which images are used (for example, in assessing the anatomy and

observations present in images when evaluating lesions in the brain).

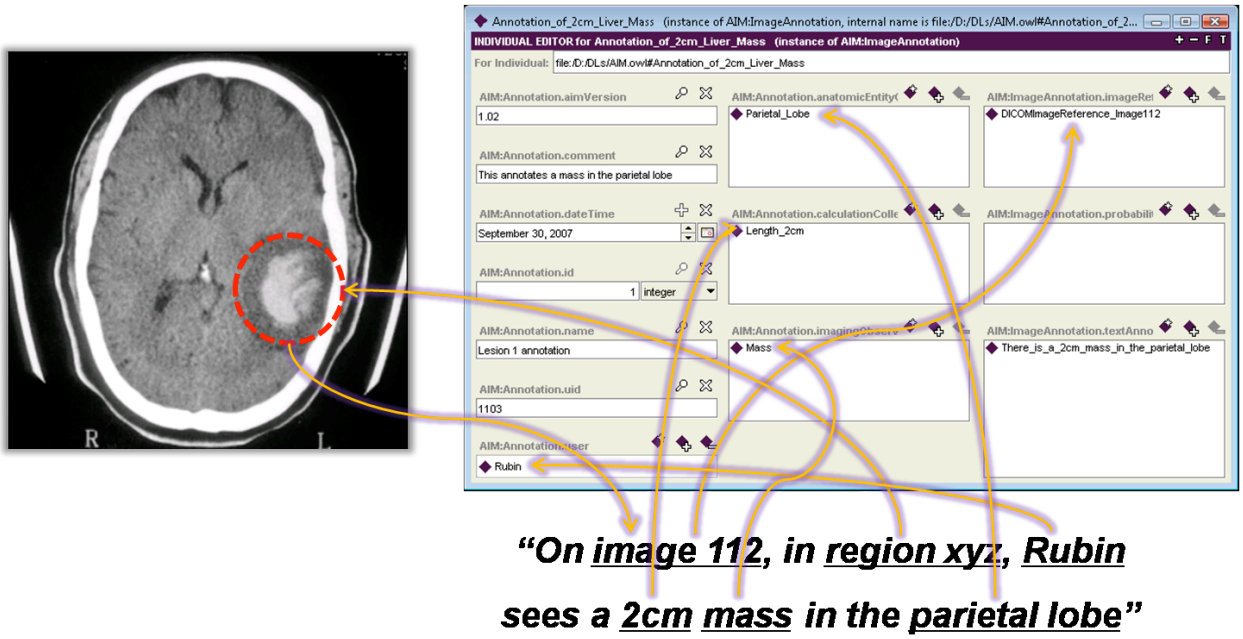


Figure 6: Semantic image annotation with AIM. AIM provides a syntax and ontology for describing the semantic content in images in a standard manner. Without AIM, workers interpreting images record their observations about images using text; in this form, the semantic image content cannot be unambiguously processed by computer applications. With AIM, the semantic content is explicit and machine-accessible, enabling applications to access this information. Accordingly, flexible queries such as "*find images showing a mass more than 1cm in size in the parietal lobe of the brain*" can be readily executed.

AIM also provides an information model ("AIM schema")—a standard syntax for creating and storing instances of image annotations. The AIM schema is in UML, and it distinguishes image "annotation" and "markup." Annotations describe the meaning in images, while markup is the visual presentation of the annotations. In the AIM schema, all annotations are either an ImageAnnotation (annotation on an image) or an AnnotationOnAnnotation (annotation on an annotation). Image annotations include information about the image as well as their semantic contents (anatomy, imaging observations, etc). Annotation on Annotations permit users to make statements about groups of pre-existing annotations, such as to comment on multi-reader image evaluations, or to make statements about a series of images.

While AIM was developed to meet the needs of cancer research, we believe that AIM will be applicable to various images in the nanotechnology domain (SEM, TEM, microscopy, *in vivo* imaging). Images acquired using nanoparticles contain rich structure and observables that are informative in nanotechnology experiments as with clinical research. We believe that AIM could enable the nanotechnology community to organize their image data more effectively and to use images for data mining for novel biomarkers.

Need for Query

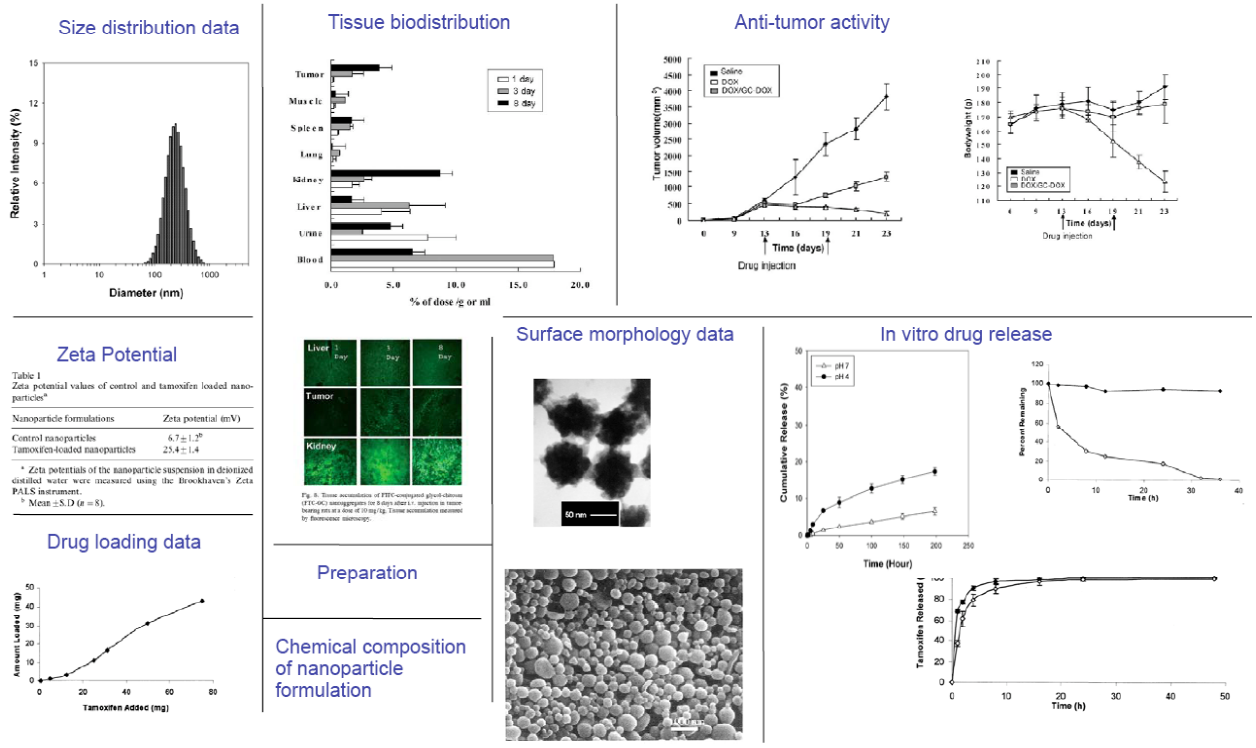
In addition to the enabling technologies described above, caBIG® is currently embarking on projects to enable data query across centers via the caGrid. Current grid query is performed using the OCQL language, which provides an SQL-like syntax for accessing data. Recently

caBIG® began embarking on a project called Image Query Formulation, which will provide semantically-based query of caBIG® image data. Queries will be formulated in terms that can be mapped to ontologies (via a "query graph"), subsequently being implemented as an ontology-based query.

Two other related projects in caBIG® will also provide means to query clinical and research data: caB2B and caIntegrator. caB2B simplifies query construction and makes it easy for non-experts and non-computational people to formulate queries in intuitive ways. caIntegrator provides a simple portal to virtual integration of a variety of clinical and experimental data and users may construct searches based on a variety of different data attributes.

5.3 Nanotechnology Is Very Integrative

Nanotechnology research generates a wide variety of data types as is exemplified in Figure 7. Additionally, many other technologies may be used such DNA microarrays, flow cytometry, magnetic sorting, microfluidics, etc. Because of this, an integrative environment similar to that provided by tools such as caIntegrator may be very useful. However, unlike some of the resources currently built with caIntegrator, there may not be the canonical type of workflows but rather a very large number of different ways in which these diverse data types are used.



Source: Son et al, J. of Controlled Release, **51**, 135-145 (2003)

Figure 7: An example from the literature of the diversity of different data types involved in a nanotechnology experiment.

5.4 Database Resource Gaps and Needs

caNanoLab is under continuing development and enhancement and represents the major

information resource for primary cancer nanotechnology data. The caNanoLab development team is extremely responsive to its user community and works proactively to address gaps and needs. User feedback is collected on an informal basis through e-mail and a user mailing list as well as through more formal monthly user group meetings. All requirements and feature requests for caNanoLab have been collected on the caNanoLab project Tracker list (http://gforge.nci.nih.gov/tracker/?atid=368&group_id=69&func=browse) and are being addressed by the caNanoLab development team.

However, caNanoLab is limited by the rate at which data can be added to the resource because it has not yet achieved widespread adoption. The top need for caNanoLab is its continued growth through new deposition of cancer nanotechnology. This growth is necessary to encourage widespread adoption of the resource: potential users need to see valuable data in caNanoLab in order to recognize the benefits of such a database. An optimal approach to caNanoLab growth will involve both (1) continued curation of the literature to provide focused entry of data in specific areas of need as well as (2) efforts to promote broader community use and adoption. Literature curation provides a steady flow of well-curated data that can be focused to specific areas of research need or interest. Additionally, caNanoLab curators can be used to "seed" caNanoLab with data from specific research groups to help encourage community "buy in". Community participation can also be encouraged through education and training activities to demonstrate the power of caNanoLab and illustrate its use for data archiving, searching, and analysis.

Computational modeling represents another approach related to nanotechnology informatics and there is considerable work already in this area. A modeling collaboratory might provide a flexible mechanism for exchanging models, model parameters, results, validation and test suites as well as annotation of these entities by outside users. Currently, the NanoHUB project provides some of this functionality for sharing models. The need exists for all modeling and simulation for nanotechnology needs, including quantum calculations, mesoscale simulations, organ and system models, SARs, PK/PB simulations, manufacturing and scale-up models, transport models, risk assessment models, etc. In addition, the structural models of the nanomaterial itself require large expenditures of resources to develop and should be shared as well. The Linnaeus Project is an example of the collaborative mechanisms needed to accelerate model development, testing, re-use, annotation, and curation.

6 Knowledge Resources

Our understanding of cancer biology has accelerated with high throughput assays in the -omics era. Translational researchers using nanotechnology need to integrate this new information into their development of new approaches for cancer diagnosis and therapy. The flood of information from challenges our cognitive capacity to recall the latest and most relevant information. The skills needed to navigate through this ocean of information can be extremely high barrier, especially given the very interdisciplinary nature of the field. Researchers need tools that can simplify and speed up the retrieval of information from a broad spectrum of changing knowledge sources.

6.1 Diverse Knowledge Sources

The scientific literature on nanoparticles in biological applications is growing exponentially,

with many entire journals now devoted to nanotechnology (e.g., *Nature Nanotechnology*, *Nanobiotechnology*, *ACS Nano*, *ACS Nano Letters*, *IEEE Transactions on Nanobioscience,* and many more). Many of the biologically relevant portions of this literature are indexed within PubMed; however, there are several more chemical, physical, and engineering aspects of this literature that are not included in PubMed. While citation indexes that span more these areas exist (e.g., SciSearch, Scopus, Web of Knowledge, Faculty of 1000, etc.), much of the infrastructure within caBIG® is built upon using PubMedIDs.

Other sources of information are relevant as well including the U.S. Patent and Trademark Office, commercial vendors (of both nanomaterials and instruments), clinical trials (e.g., access to what trials have been terminated and why), FDA, environmental health and safety databases, etc. While each database alone is functional, there is a need for re-factored access to information across these databases in order to simplify and expedite information retrieval in this area. As mentioned elsewhere, there is a need for structured access to this information although this presents a daunting task since these various sources do not adhere to any common data-structuring scheme.

6.2 Search

Additionally, literature searching for cancer nanotechnology research articles is further complicated by the lack of a standardized nomenclature for nanoparticles; e.g., analogous to IUPAC nomenclature for chemical compounds. Such standardized nomenclature does not exist, thus creating synonymy problems that are compounded by the interdisciplinary nature of cancer nanotechnology research. For instance, the terms "quantum dot", "q-dot", "QD" and "semiconductor nanocrystal" are interchangeable; they all refer to the same class of nanoparticle. A clinical researcher searching for information on these types of nanoparticle platforms would require *a priori* knowledge of all of these terms in order to perform a comprehensive literature search. Although some of these terms exist in the National Library of Medicine's MeSH vocabulary, not all of these synonyms are entered and the number of new terms for new classes of nanoparticles is expanding extremely rapidly. Furthermore, different types of nanoparticles can have distinct axes of similarity depending on their composition. For instance, some sets of nanoparticles may share a core material in common while other sets of nanoparticles may share targeting moieties in common. These axes are not captured well without a structured and rapidly updated vocabulary for clearly describing concepts related to nanoparticle composition. All of these vocabulary requirements strongly indicate the need for an ontology-driven approach to the description of cancer nanotechnology research literature and knowledge. Such an approach is discussed in the next section.

6.3 Knowledge Resource Gaps and Needs

A Knowledge Portal

The cancer nanotechnology community needs a centralized portal for knowledge sources that spans the full range of disciplines represented in this very diverse field. Ideally, this resource would be a superset of PubMed with the same straightforward interface and query syntax that users are already familiar with. Additionally, links between this resource and other NIH databases (including chemical, nanoparticle/caNanoLab, etc.) would be extremely beneficial. The caNanoLab development team has already provided infrastructure to support "linking" of nanoparticle entries with PubMed articles, so the groundwork for such integration is already

available.

Semantically Aware Literature Search and Beyond

Related to this, the ability to retrieve remote information dispersed throughout many papers would be powerful and is sorely needed. For example, if a researcher would like to see information on the blood half-life of all gold-based nanomaterials regardless of shape and size, there is no way to do this without going through multiple searches, and retrieving the publications and collating the relevant information from each publication. Similar problems exist for gaining easy access to the figures and tables within each publication. For example, retrieving dose response curves from all recent articles on a specific nanoparticle would currently have to be done manually. An effort to address these issues would likely require a dedicated curation staff with long term programmatic support. An additional useful feature would be a tool that constructs complex literature queries with the ability to save the protocols for running at a later time or by other researchers and remote collaborators (in a similar manner to caB2B or caIntegrator).

Communication Across Fields

Nanotechnology development for medical applications sits in a chasm between material scientists and clinical physicians. Effective communication between the widely disparate fields is a challenge that informatics can address most efficiently. Bioinformatics can provide a central location for communication across fields, like a knowledgebase for providing some basic guidelines for cancer applications. A material scientist may spend considerable time in a direction with the 'long-term goal' of cancer application. However, because he/she is so far removed from feasible biological applications, efficiency through communication with others in the community can forge new collaborations and accelerate progress. Clearly, the problem here is competition. Open forums for communication, similar to a wiki can provide a date/time log and also introduce corrective actions against bad information.

7 Recommendations

This section contains the recommendations of the caBIG® ICR Nanotechnology Informatics Working Group that have resulted from several months of discussions, including meetings at both the caBIG® Annual Meeting in Washington DC in June 2008 and the caBIG® ICR Face-to-Face in Boston in September 2008. Additionally, we have received invaluable feedback from other members of the nanotechnology community.

These recommendations represent high value areas that are immediately actionable and could provide the greatest impact for both the caBIG® community as well as the nanotechnology community.

In the following sections, asterisks preceding a recommendation denote high priority areas that the working group identified as near-term goals for completion.

7.1 Formation of a Global Community of Interest

The new NanoCollaboratory wiki is being considered as the primary mechanism to communicate for the IANH. A similar effort should be undertaken to unite the vocabulary and ontology development efforts. A special effort should be made to integrate professional

organizations into this new alliance. The NLM, ACS, APS could all sponsor authoritative terminology and definitions for this effort, although obstacles such as intellectual property much be addressed early on. caBIG® could function as a focus for this effort, especially due to its existing framework and open software. The NPO could serve as an initial (English) ontology, and cross-linking to other nanotechnology ontologies would be a substantive, consensus-building project. The existence of this community of interest would greatly enhance the other recommendations and allow both top-down and bottom-up development of the ontology due to the existing consensus regarding international collaboration and resource sharing in standards development.

7.2 Vocabulary and Ontology

As a basic tenet of caBIG® and biomedical informatics in general, information in the area of cancer nanotechnology research should be formally encoded using standardized vocabularies and structured ontologies to enable effective sharing of information [50]. Such an ontology should ideally be machine parseable as well as human readable.

The continued development of an ontology to cover the cancer nanotechnology domain is a critical effort toward this end. To our knowledge, the Nanoparticle Ontology (NPO) developed by the Baker and Pappu laboratories at Washington University is the only extensive ontology in this domain, and its public release is imminent. The NPO can thus serve as the basis for ongoing efforts to develop a comprehensive and dynamic ontology that catalyzes nanomaterial applications to cancer. Recommendations towards this end include:

a. ** Formal integration into the caBIG® framework. For instance, on its way to achieving caBIG® Silver level compatibility, the caNanoLab tool should adopt this ontology into its domain object model. Similarly, NPO should be compliant with the standards established by VCDE to achieve certification as a caBIG® standardized ontology.

b. ** Planning for the long-term funding, maintenance and governance of NPO, such as through a consortium of stakeholders in academics, industry, government, and private foundations.

c. Integration with existing ontologies and databases that are relevant to nanomaterial composition or their biomedical applications, such as ChEBI, PubChem and ChemBank for small molecules; Gene Ontology (GO) for protein targets of nanomaterials; and clinical trials databases such as clinicaltrials.gov.

d. Development of a structured mechanism for open feedback and evolution of NPO by the community. Currently, the ontology is available through BioPortal (http://bioportal.bioontology.org/) as proof-of-concept. Continued involvement from a wider swath of domain experts in the nanotechnology community will greatly advance both the continued development of the ontology (e.g., to incorporate emerging categories such as manufacturing processes, or biologic activity) as well as its adoption by the research community.

7.3 Minimum Information Standards

Minimum information standards such as the Minimum Information About a Microarray Experiment (MIAME) standard have become essential for allowing data from diverse laboratories to be collected and shared to allow the understanding of the context, methods and

data of an experiment [49]. Currently, no such standard or checklist exists for experiments in the cancer nanotechnology domain. We propose that such a standard should be formulated and collaboratively developed amongst the major stakeholders in cancer nanotechnology.

a. ** A standard for minimum information about a nanoparticle experiment would describe nanoparticle synthesis, nanoparticle composition and topology, physical and chemical properties (e.g., size, shape, state of dispersion), best practices for characterization, surface coatings and properties, functional ligands, remote sensing properties, and intended modality of use (*ex vivo* sensing, *in vivo* imaging, therapeutic, theranostic). Minimum data should also include data on the properties of the nanomaterial in cell culture or animal model systems (including human studies where applicable), including tissue distribution, pharmacokinetics, biological activity (*in vitro* and *in vivo*) including toxicity and adverse side effects, and metadata for characterization of biological activity. The standard could also eventually include a set of optimal assays that could contribute to a minimal biological characterization of nanomaterials.

b. ** Build upon, and/or integrate with existing efforts such as caNanoLab and the Minimum Information for Biological and Biomedical Investigations (MIBBI) [51].

c. Encourage awareness and adoption of these standards by the community, e.g., by linking deposition of formatted data and metadata as part of the journal publication process or submission for FDA approval.

d. Establish a mechanism for the long-term maintenance and governance of such a standards resource. This should be coordinated with existing efforts by the SDOs and OECD mentioned above

7.4 Dense Data Generation

A major issue with the current state of nanotechnology informatics is the dearth of database entries. caNanoLab is an existing nanoparticle database that has been available to the public for approximately one year. A designated curator extracts nanocharacterization data from papers arising from all eight CCNE centers. (Similarly, curators manually populate the Molecular Imaging and Contrast Agent Database (MICAD) with properties of various imaging agents, including numerous nanoparticle-based imaging agents.) Despite these resources, the field of nanomaterials remains very "data poor" in comparison to microarray data, for example. This, in turn, limits the ability to discern structure-activity relationships, or similarities or differences among different materials. There are many reasons for the relative dearth of data, including the existence of multiple, inadequately publicized repositories, the reliance on post hoc curation, and hesitation about data-sharing (for understandable competitive or intellectual property reasons).

Systematic studies have produced a dense matrix of data (measurements × nanoparticles) for as many as 50 nanoparticles, which enabled the discovery of structure-activity relationships [52]. Future systematic studies could produce data matrices that when quilted together will encompass the space of nanomaterial compositions, properties, and biologic activities. This would require standardization of the experiments themselves (both which experiments to perform, and the experimental protocols) and widespread acceptance of the descriptors of these experiments. Although this space may never be exhaustively mapped due to its exponential size [14], the systematic mapping of subsets of this space will enable secondary usage of this

data for structure-activity studies, and in the future, aid in the de novo design of biologically inert and safe nanomaterials. Such dense mapping can be encouraged by the following activities:

a. ** Working towards a standardized suite of *in vitro* assays to annotate the biological activity of nanoparticles, a near-term goal would be to systematically characterize a collection of representative benchmark nanomaterials. The goal would be to extend the approaches described in two systematic studies [52, 53] to a broader variety of materials, and test the feasibility and utility of such approaches

   - Selection of nanomaterials would be based on several criteria: (a) structural diversity (e.g., based on a wide variety of nanomaterial platforms, size, shape, topology, composition); (b) capable of physical and chemical characterization in accordance with minimum information standards; (c) existence of pre-existing or newly generated in vivo data in animal models or humans, such as biodistribution, pharmacokinetics, biologic activity, toxicity (or relative safety, as defined by lack of adverse effects in human trials); (d) availability in sufficient quantities (by standardized, reproducible synthetic methods) for systematic screening, (e) the availability of a structural model providing a description of the representative molecular structures and conformations present in the polydisperse population of structures in a typical sample.
   - Systematic assessment of the effects of these nanomaterials on a collection of diverse cell types, using multiple assay measurements. Measurements could reflect potential mechanisms of cytotoxicity (apoptosis, impaired mitochondrial function, generation of reactive oxygen species, etc.) and/or aspects of cellular physiology (growth rate, DNA replication, protein synthesis, etc.). In addition, systematic morphologic assessments could be made using automated high-content imaging screens [53].
   - Analysis of the resulting data to address several proof-of-concept questions: To what extent are different assay measurements orthogonal? Is there a subset of measurements that confers the same information as the entire list of measurements, and that could serve as a core set of benchmarking assays? To what extent are structure-activity relationships discernable across widely divergent nanomaterial platforms?

b. ** Establishment or adoption of a central repository such as caNanoLab for characterization and biological activity data on nanomaterials, perhaps by expanding existing efforts as part of CCNEs. Minimum information standards would make formatting metadata for upload more uniform. This data repository would ideally also include relevant analytic tools as described below (Section 7.5).

c. Encouraging an ethos of data sharing in the interest of advancing the pace of the community's overall research efforts. For instance, contributions of data to a central repository could be encouraged by several steps: (a) following a quarantine period, contributors would be able to access data on materials contributed by other investigators; (b) in exchange for donating materials, "core" services could be made available to perform standard characterization studies, and upload data to the repository; (c) availability of analytic tools as part of the data repository; (d) clarification and reasonable protection of intellectual property interests surrounding

contributed data; (e) requirements that submission of data must accompany manuscript publication in participating journals.

7.5 Analytical Tools and Services

The final area of recommendations concerns analytical tools that will enhance the functionality of nanotechnology databases such as caNanoLab. These include methods for the calculation and prediction of structure-activity relationships, nanotechnology-related information and image repositories, and tools for nanotechnology characterization. These tools should be integrated into emerging standardized nanotechnology databases, to create data repositories and analytic environments that can catalyze research. Furthermore, the use of certain more powerful analysis tools could be reserved for those users who have contributed data and agreed to data-sharing principles. Finally, additional sophisticated tools for the analysis of generic large datasets are already available through caBIG®, and the implementation of standard data formats and minimum information standards will facilitate the application of these methods to the analysis of nanotechnology-related datasets.

a. ** Ensure that emerging nanotechnology tools (e.g., caNanoLab) are compatible with caBIG® standards and integrated into existing caBIG® tools (such as caTissue, caArray, GenePattern, geWorkBench, NCIA, etc.) and packages (e.g., Life Sciences Distribution or caIntegrator2).

b. ** Create a publicly accessible information portal for a variety of nanotechnology-related information, such as literature citations, clinical trial updates, patent databases, and image repositories. Because a major application of nanotechnology lies in molecular imaging, a searchable database of published images could benefit a wide community. (An example from radiology, the ARRS GoldMiner project, indexes publication figures based on keywords in the figure legends [54])

c. Development and dissemination of tools for quantitative structure-activity relationships among nanomaterials. These tools should be predictive, and generate hypotheses that can be tested on independent datasets or with independent experiments.

d. Integrate analytic tools as they become available into a community-wide nanotechnology data repository and analysis environment. The availability of these tools would provide further incentive for investigators to contribute data, and participate in data-sharing agreements.

e. Development and disseminate tools to characterize physical properties of the nanoparticles themselves. (For instance, TEM and SEM images may be automatically analyzed to determine the distribution of shapes and sizes of nanoparticles. Spectral optical data such as from Surface Enhanced Raman Spectroscopy (SERS) nanoparticles may require specialized algorithms for spectral unmixing.)

f. Coordinate some effort to use the ontology in searches over the Nano Linnaeus Project to supplement experimental characterization with modeling data. The NCL has several examples of these calculations, which could be used as examples. In addition, there are existing organ models relevant to PK/PD modeling, as well as meso–scale models for modeling efficacy (e.g. NIST models for heating of gold nanoparticles through the plasmon resonance).

8 Path Toward the Future

NCI has set an ambitious goal of "ending suffering from cancer" by the year 2015. In addition to this challenging target, the mandate also calls for personalized cures for cancer. Personalized medicine involves the prescription of therapeutics best suited for a given individual. Such customized therapeutics will be designed based on individual pharmacogenetic, pharamacogenomic, and pharmacoproteomic information, as well as information regarding the behavior of cancer in a given individual. To achieve the goals of personalized medicine, clinicians will increasingly rely on high-throughput and high-precision data gathered using nanodiagnostics as well as predictions of disease progression based on quantitative models that are driven by biomarker and other molecular-level data, which in all likelihood will also be gathered using the tools from nanotechnology. The challenges of personalized medicine are best captured in the following opinion, extracted verbatim from K.K. Jain's recent review on advances in nanooncology:

> "With so many nanotechnologies available for drug delivery, it is recommended that computational and mathematical tools be used to predict which nanovectors, surface modifications, therapeutic agents, and penetration enhancers to use for a multi-stage drug delivery strategy that would enable efficient localized delivery of chemotherapeutic drugs leading to significant improvements in therapy efficacy and reduced systemic toxicity. Such an approach can be optimized for personalized oncology" [55].

We see a tremendous opportunity for informatics to accelerate discovery and translation of cancer nanotechnology. In order to realize this vision, however, there must be an investment into further development of the informatics infrastructure that will drive progress in this area.

9 References

1. Feynman, R., *There's Plenty of Room at the Bottom.* Talk delivered at American Physical Society Meeting, 1959.
2. Roco, M.C., *Nanotechnology: convergence with modern biology and medicine.* Curr Opin Biotechnol, 2003. **14**(3): p. 337-46.
3. Whitesides, G.M., *The 'right' size in nanobiotechnology.* Nat Biotechnol, 2003. **21**(10): p. 1161-5.
4. Choi, H.S., et al., *Renal clearance of quantum dots.* Nat Biotechnol, 2007. **25**(10): p. 1165-70.
5. Euliss, L.E., et al., *Imparting size, shape, and composition control of materials for nanomedicine.* Chem Soc Rev, 2006. **35**(11): p. 1095-104.
6. Byrne, J.D., T. Betancourt, and L. Brannon-Peppas, *Active targeting schemes for nanoparticle systems in cancer therapeutics.* Adv Drug Deliv Rev, 2008.
7. McCarthy, J.R. and R. Weissleder, *Multifunctional magnetic nanoparticles for targeted imaging and therapy.* Adv Drug Deliv Rev, 2008. **60**(11): p. 1241-51.
8. Wang, A.Z., et al., *Biofunctionalized targeted nanoparticles for therapeutic applications.* Expert Opin Biol Ther, 2008. **8**(8): p. 1063-70.
9. Hawker, C.J. and K.L. Wooley, *The convergence of synthetic organic and polymer chemistries.* Science, 2005. **309**(5738): p. 1200-5.
10. Wolinsky, J.B. and M.W. Grinstaff, *Therapeutic and diagnostic applications of dendrimers for cancer treatment.* Adv Drug Deliv Rev, 2008. **60**(9): p. 1037-55.
11. Perez, J.M., et al., *Magnetic relaxation switches capable of sensing molecular interactions.* Nat Biotechnol, 2002. **20**(8): p. 816-20.
12. Verma, A., et al., *Surface-structure-regulated cell-membrane penetration by monolayer-protected nanoparticles.* Nat Mater, 2008. **7**(7): p. 588-95.
13. Ferrari, M., *Cancer nanotechnology: opportunities and challenges.* Nat Rev Cancer, 2005. **5**(3): p. 161-71.
14. Ferrari, M., *The mathematical engines of nanomedicine.* Small, 2008. **4**(1): p. 20-5.
15. Grodzinski, P., M. Silver, and L.K. Molnar, *Nanotechnology for cancer diagnostics: promises and challenges.* Expert Rev Mol Diagn, 2006. **6**(3): p. 307-18.
16. Pierce, M.C., D.J. Javier, and R. Richards-Kortum, *Optical contrast agents and imaging systems for detection and diagnosis of cancer.* Int J Cancer, 2008. **123**(9): p. 1979-90.
17. Weissleder, R. and M.J. Pittet, *Imaging in the era of molecular oncology.* Nature, 2008. **452**(7187): p. 580-9.
18. Xing, Y., et al., *Molecular profiling of single cancer cells and clinical tissue specimens with semiconductor quantum dots.* Int J Nanomedicine, 2006. **1**(4): p. 473-81.
19. Lee, H., et al., *Chip-NMR biosensor for detection and molecular analysis of cells.* Nat Med, 2008. **14**(8): p. 869-74.

20. Peer, D., et al., *Nanocarriers as an emerging platform for cancer therapy.* Nat Nanotechnol, 2007. **2**(12): p. 751-60.

21. Harisinghani, M.G., et al., *Noninvasive detection of clinically occult lymph-node metastases in prostate cancer.* N Engl J Med, 2003. **348**(25): p. 2491-9.

22. Gradishar, W.J., *Albumin-bound paclitaxel: a next-generation taxane.* Expert Opin Pharmacother, 2006. **7**(8): p. 1041-53.

23. Rupp, R., S.L. Rosenthal, and L.R. Stanberry, *VivaGel (SPL7013 Gel): a candidate dendrimer--microbicide for the prevention of HIV and HSV infection.* Int J Nanomedicine, 2007. **2**(4): p. 561-6.

24. Campbell, S., et al., *Clinical applications of a direct genomic platform for SNP detection in human DNA samples using gold nanoparticle probe technology and enhanced signal amplification.* Journal of Thrombosis and Haemostasis, 2007. **5**(Supplement 1): p. P-W-677.

25. Maheswaran, S., et al., *Detection of mutations in EGFR in circulating lung-cancer cells.* N Engl J Med, 2008. **359**(4): p. 366-77.

26. Nagrath, S., et al., *Isolation of rare circulating tumour cells in cancer patients by microchip technology.* Nature, 2007. **450**(7173): p. 1235-9.

27. Stern, S.T. and S.E. McNeil, *Nanotechnology safety concerns revisited.* Toxicol Sci, 2008. **101**(1): p. 4-21.

28. Maynard, A.D., et al., *Safe handling of nanotechnology.* Nature, 2006. **444**(7117): p. 267-9.

29. Committee to Review the National Nanotechnology Initiative, N.R.C., *A matter of size: Triennial review of the National Nanotechnology Initiative (NMAB)*. 2006, Washington, DC: The National Academies Press.

30. The Nanotechnology Task Force, *Nanotechnology: A Report of the U.S. Food and Drug Administration Nanotechnology Task Force*. 2007, U.S. FDA.

31. DeAngelis, C.D., et al., *Clinical trial registration: a statement from the International Committee of Medical Journal Editors.* JAMA, 2004. **292**(11): p. 1363-4.

32. Fragoso, G., et al., *Overview and Utilization of the NCI Thesaurus.* Comp Funct Genomics, 2004. **5**(8): p. 648-54.

33. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

34. Degtyarenko, K., et al., *ChEBI: a database and ontology for chemical entities of biological interest.* Nucleic Acids Res, 2008. **36**(Database issue): p. D344-50.

35. Musen, M.A., *Dimensions of knowledge sharing and reuse.* Comput Biomed Res, 1992. **25**(5): p. 435-67.

36. Gruber, T.R., *A translation approach to portable ontology specifications.* Knowledge Acquisition, 1993. **5**(2): p. 199-220.

37. Stevens, R., C.A. Goble, and S. Bechhofer, *Ontology-based knowledge representation for bioinformatics.* Brief Bioinform, 2000. **1**(4): p. 398-414.

38. Grenon, P., B. Smith, and L. Goldberg, *Biodynamic ontology: applying BFO in the*

biomedical domain. Stud Health Technol Inform, 2004. **102**: p. 20-38.

39. Cheung, K., J. Drennan, and J. Hunter, *Towards an Ontology for Data-driven Discovery of New Materials.* AAAI Spring Symposium, Semantic Scientific Knowledge Integration, 2008.
40. Hunter, J., *Scientific Models: A User-oriented Approach to the Integration of Scientific Data and Digital Libraries.* VALA2006, 2006.
41. Kozaki, K., Y. Kitamura, and R. Mizoguchi, *Systematization of Nanotechnology Knowledge Through Ontology Engineering -- A Trial Development of Idea Creation Support System for Materials Design based on Functional Ontology.* Proceedings of International Semantic Web Conference, 2003.
42. Kozaki, K., Y. Kitamura, and R. Mizoguchi, *Developing Ontology-based Applications using Hozo.* Proceedings of the Fourth IASTED International Conference on Computational Intelligence, 2005.
43. *http://mandala.t.u-tokyo.ac.jp/english/nanoindex.html*. [cited.
44. Tanaka, M., *Toward a Proposed Ontology for Nanoscience.* Proceedings of Canadian Association for Information Science, 2005.
45. Gilliam, L., M. Tariq, and K. Ahmad, *Terminology and the construction of ontology.* Terminology, 2005. **11**(1): p. 55-81.
46. Niemann, B., *Nanoinformatics: Locate, Collaborate, and Integrate.* Workshop on Nanoinformatics Strategies, 2007.
47. *http://www.nanosense.org/documents/nanoed05/AtlasOfNanotechnology.pdf*. [cited.
48. *http://uqbar.ncifcrf.gov/groups/nlinnaeus/*. [cited.
49. Brazma, A., et al., *Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.* Nat Genet, 2001. **29**(4): p. 365-71.
50. von Eschenbach, A. and K. Buetow, *Cancer Informatics Vision: caBIG.* Cancer Informatics, 2006. **2**: p. 22-24.
51. Taylor, C.F., et al., *Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project.* Nat Biotechnol, 2008. **26**(8): p. 889-96.
52. Shaw, S.Y., et al., *Perturbational profiling of nanomaterial biologic activity.* Proc Natl Acad Sci U S A, 2008. **105**(21): p. 7387-92.
53. Jan, E., et al., *High-Content Screening as a Universal Tool for Fingerprinting of Cytotoxicity of Nanoparticles.* ACS Nano, 2008. **2**(5): p. 928-38.
54. Kahn, C.E., Jr. and C. Thao, *GoldMiner: a radiology image search engine.* AJR Am J Roentgenol, 2007. **188**(6): p. 1475-8.
55. Jain, K.K., *Recent advances in nanooncology.* Technology in Cancer Research and Treatment, 2008. **7**(1): p. 1-13.